Abstract

After trying Festival Speech Synthesis System, we decided to develop our own TTS framework, conducive to perform the necessary research experiments for developing good quality TTS for Indian languages. In most of the attempts on Indian language TTS, there is no prosody model, provision for handling foreign language words and no phrase break prediction leading to the possibility of introducing appropriate pauses in the synthesized speech. Further, in the Indian context, there is a real felt need for a bilingual TTS, involving English, along with the Indian language. In fact, it may be desirable to also have a trilingual TTS, which can also take care of the language of the neighboring state or Hindi, in addition. Thus, there is a felt need for a full-fledged TTS development framework, which lends itself for experimentation involving all the above issues and more.

This thesis work is therefore such a serious attempt to develop a modular, unit selection based TTS framework. The developed system has been tested for its effectiveness to create intelligible speech in Tamil and Kannada. The created system has also been used to carry out two research experiments on TTS.

The first part of the work is the design and development of corpus-based concatenative Tamil speech synthesizer in Matlab and C. A synthesis database has been created with 1027 phonetically rich, pre-recorded sentences, segmented at the phone level. From the sentence to be synthesized, specifications of the required target units are predicted. During synthesis, database units are selected that best match the target specification according to a distance metric and a concatenation quality metric. To accelerate matching, the features of the end frames of the database units have been pre-computed and stored. The selected units are concatenated to produce synthetic speech. The high values of the obtained mean opinion scores for the TTS output reveal that speech synthesized using our TTS is intelligible and acceptably natural and can possibly be put to commercial use with some additional features. Experiments carried out by others using my TTS framework have shown that, whenever the required phonetic context is not available in the synthesis database., similar phones that are perceptually indistinguishable may be substituted.

The second part of the work deals with the design and modification of the developed TTS framework to be embedded in mobile phones. Commercial GSM FR, EFR and AMR speech codecs are used for compressing our synthesis database. Perception experiments reveal that speech synthesized using a highly compressed database is reasonably natural. This holds promise in the future to read SMSs and emails on mobile phones in Indian languages. Finally, we observe that incorporating prosody and pause models for Indian language TTS would further enhance the quality of the synthetic speech. These are some of the potential, unexplored areas ahead, for research in speech synthesis in Indian languages.